A matched-filter technique with an objective threshold

Shiro Hirano[1], Hironori Kawakata[1], and Issei Doi[2]

[1]Department of Physical Science, College of Science and Engineering, Ritsumeikan University, 1-1-1, Nojihigashi, Kusatsu, Shiga, 525-8577, Japan.

[2]Disaster Prevention Research Institute, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan.

Key Points:

- We developed an objective thresholding method for seismic event detection.
- We show that cross-correlation coefficients among waveforms do not necessarily follow a normal distribution.
- Even in the case, we can detect ouliers reasonably and objectively on the basis of Akaike's information criterion and extreme value theory.

Index Terms

7230 Seismicity and tectonics

Keywords

- Seismicity
- Seismic event detection
- Waveform cross-correlation
- Mathed-filter analysis
- Information criterion
- Extreme value statistics

Corresponding author: Shiro Hirano, `s-hrn@fc.ritsumei.ac.jp`

Abstract

We propose an objective threshold determination method for detecting outliers from the empirical distribution of cross-correlation coefficients among seismic waveforms. This method is aimed at detecting seismic signals from continuous waveform records. In our framework, detectability is automatically determined from Akaike's Information Criterion (AIC). We applied the method of seismic signal detection to continuous records collected over 2 years. The results show that the maximum value of network cross-correlation coefficients sampled from each constant interval can be approximated by the theory of extreme value statistics, which provides a parametric probability density function of maxima. Using the function, outliers can be considered with a reasonable criterion.

1 Introduction

A matched-filter (MF) analysis, which is a technique for quantifying the similarity between continuous and template waveforms using the cross-correlation coefficient (CC), is efficient in detecting weak seismic signals embedded in continuous waveform records [Gibbons & Ringdal, 2006]. Many types of seismic events have been detected automatically using MF analysis: non-volcanic tremors and low frequency earthquakes [e.g., Shelly et al., 2007; Ohta & Ide, 2008; Aso et al., 2011], seismic swarms [e.g., Shimojo et al., 2014; Ohmi, 2015], and foreshocks and aftershocks [e.g., Bouchon et al., 2011; Kato et al., 2012; Doi & Kawakata, 2012, 2013]. In general MF analyses, waveforms are regarded as seismic signals when the CC between a template and continuous waveform exceeds a threshold value. The threshold value has occasionally been defined as a constant [Doi & Kawakata, 2012, 2013] or not specified [Bouchon et al., 2011]. However, given the possibility of relatively high CC values randomly occurring for microtremors, the threshold should be determined depending on the empirical frequency distributions of CC. In other previous studies, the threshold value was defined as a constant factor multiplied by the standard deviation (σ) [Ohta & Ide, 2008; Aso et al., 2011] or the median absolute deviation (MAD) [Shelly et al., 2007]. Under this strategy, we can estimate the possibility of a false positive if a probability density function (PDF) of the CC is known. Thus, characteristics of the PDF should be investigated both theoretically and experimentally. Because event detection is a type of outlier detection, careful attention should

be given to the tails of the frequency distribution of CC; do they follow the Gaussian, exponential, or power law? Only Aso et al. [2011] showed that the tail follows a normal distribution in their case. In this study, we first derive a normal distribution that the CC between random microtremor and random template waveform may follow and investigate the effect of a band-pass filter, which provides reference for determining a realistic CC distribution. Next, we consider a distribution that the maximum value of CC in every constant interval follows for robust outlier detection using non-random continuous waveform records. The distribution of maxima in every constant interval is given by the extreme value theory [Gumbel, 1958]. Subsequently, we demonstrate that the tails of CC values are well modeled by the extreme value distribution rather than the normal distribution through a case study of 2-years continuous records and multiple templates of foreshocks before an M5.4 crustal earthquake in Japan. Given the extreme value distribution, we employ a reasonable method for detecting outliers based on Akaike's Information Criterion (AIC). Although we focus on a specific foreshock activity in our data analysis section, the method proposed in this study is efficient for other seismic phenomena and regions.

2 Theory and Method

2.1 Ideal frequency distribution of CC

In the following, without loss of generality, we regard velocity seismograms as the data. The frequency distribution of CC between a continuous record and an arbitrary template waveform array of length d follows a normal distribution whose mean is zero and variance is d^{-1} if the continuous record is an independent and identically distributed (i.i.d.) random variable. Let d-dimension vectors $\boldsymbol{u} := (u_i)$ and $\boldsymbol{v} := (v_i)$ $(i = 1, 2, \ldots, d)$ be discretized and offset-eliminated waveform arrays of length d. Their CC is given as

$$CC = \widehat{\boldsymbol{u}} \cdot \widehat{\boldsymbol{v}}, \tag{1}$$

where $\widehat{\boldsymbol{u}}$ and $\widehat{\boldsymbol{v}}$ are normalized $\boldsymbol{u}$ and $\boldsymbol{v}$, respectively, by their L_2-norm. If $\boldsymbol{v}$ is extracted from a random waveform, the normalized vector $\widehat{\boldsymbol{v}}$ is an isotropic random vector restricted on the $(d-1)$-dimensional unit sphere. Because eq.(1) is a projection of $\widehat{\boldsymbol{v}}$ along the $\widehat{\boldsymbol{u}}$-direction, CC can be regarded as a velocity component along the $\widehat{\boldsymbol{u}}$-direction of randomly hurtling particles with unit velocity ($|\widehat{\boldsymbol{v}}| \equiv 1$). There-

fore, the PDF of CC can be approximated by extending the Maxwell-Boltzmann distribution from 3-dimensional space to d-dimensional space; see also Appendix A. In fact, the template and continuous waveform are filtered in MF analyses because seismic waveforms have high S/N ratios in some limited frequency bands. Linear band-passed filtering is equivalent to the convolution of a characteristic function and the continuous waveform, and therefore, $\boldsymbol{v}$ is not a complete random vector but necessarily has interdependence among some neighbor samples (referred to as "self-interdependence") depending on the filter. Thus, we conducted numerical experiments; we calculate CC between an i.i.d. random waveform of length 10^8 and a random array of length $d = 500$ (figure 1). After 10 experiments, we confirmed that CC follows the normal distribution $\mathcal{N}\left(0, d^{-1}\right)$ as expected above and in Appendix A. On the other hand, if we regard the waveforms as 100 Hz time-series and apply a band-pass filter of $5-30$ Hz that is required in the next section, we find that the distribution is approximated as $\mathcal{N}\left(0, 1.8\, d^{-1}\right)$, as shown in figure 1. Therefore, we can conclude that CC follows the normal distribution even after applying the band-pass filter if the continuous waveform is random.

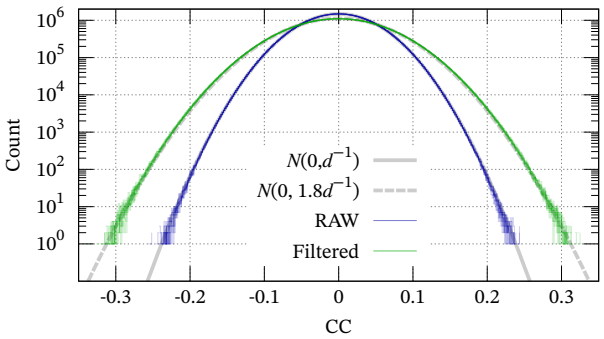

Figure 1. Frequency distribution of CC in a numerical experiment. CC between a raw random noise vs. a random template follows the normal distribution $\mathcal{N}\left(0, d^{-1}\right)$, whereas CC between a filtered random noise vs. the random template follows $\mathcal{N}\left(0, 1.8\, d^{-1}\right)$.

2.2 Frequency distribution of the maximum of CC

The assumption of i.i.d. in the previous subsection might not be valid in cases where multiple similar earthquake events frequently occur, which radiates waveforms similar to the templates, or the microtremor repeats similar patterns. In such cases,

even accidentally, relatively high CC values appear around their local peaks because of the self-interdependence. Hence, the frequency distribution of all values of CC will be contaminated by the high values repeatedly, thus rendering the tail of the distribution wider and the interpretation more difficult. To avoid this problem, outliers should be detected from the maximum value of CC in every M sample by assuming that the self-interdependence of microtremors or seismic waveforms is lost within M samples. This assumption is valid because, in general, shorter-term correlation is stronger than longer-term correlation. Theoretically, it has been shown that the frequency of the maxima of any distribution in every interval follows the Generalized Extreme Value (GEV) distribution [Gumbel, 1958; Coles, 2001], which has been employed to model possibilities of rare events, such as floods and economic crisis. GEV has the following cumulative density function (CDF):

$$F_{\mathrm{GEV}}\left(x \mid \mu', \sigma', k\right) = \exp\left(-\left(1 + k\frac{(x-\mu')}{\sigma'}\right)^{-1/k}\right), \tag{2}$$

where x is a random variable, and μ', σ' and k are the location parameter, scale parameter, and shape parameter, respectively. We have to note that $\mathrm{sgn}(k)\left(x - \mu' + \sigma'/k\right) \geq 0$ must be satisfied; otherwise the possibility is defined to be zero. It may be possible to detect outliers by fitting the distribution of the maxima with GEV even if CC does not follow the normal distribution while their maxima follow GEV; see Appendix C for the maximum likelihood estimation (MLE) of the GEV parameters. In particular, if every interval contains a sufficient amount of data, the cumulative distribution converges to one of 3 specific cases depending on the shape of their tail: the Gumbel distribution, Fréchet distribution, or Weibull distribution. In the next section, we assume that they can be approximated by the CDF of the Gumbel distribution:

$$\begin{aligned} F_{\mathrm{G}}\left(x \mid \mu', \sigma'\right) &:= \lim_{k\to 0} F_{\mathrm{GEV}}\left(x \mid \mu', \sigma', k\right) \\ &= \exp\left(-\exp\left(-\frac{x-\mu'}{\sigma'}\right)\right). \end{aligned} \tag{3}$$

This is because of the following reasons: 1) as confirmed in the next section, the accumulated data distribution shows straight falloff in semi-log plots, which is a characteristic of the Gumbel distribution, and 2) as in Appendix C, MLE of three parameters for GEV is technically difficult in some case. We focus on and plot $1 - F_{\mathrm{G}}$ in the following.

2.3 Method for event detection

Although the threshold for MF analyses has widely been assumed from the histogram of data, we have no unified or objective algorithm to assume an appropriate value of the threshold. Here, we propose an algorithm for detecting outliers reasonably and objectively on the basis of an information criterion. The elimination of outliers for minimizing AIC has been developed in applied statistics [Kitagawa, 1979; Ueda, 1996, 2009; Marmolejo-Ramos et al., 2015] and implemented in bioinformatics [Kadota et al., 2003]. Kitagawa [1979], Ueda [1996, 2009], and Kadota et al. [2003] assumed that the random variable other than the outliers follows a normal distribution and calculated AIC; Marmolejo-Ramos et al. [2015] investigated the applicability of the method in non-Gaussian and skewed distribution cases. We assume the Gumbel distribution and calculate the difference in AIC when we increase the number of suspects, which indicates whether the increment of the number is reasonable.

We sort N data in the descending order ($x_1 > x_2 > \cdots > x_N$) and assume that the leading s data $(x_1, x_2, \ldots, x_s)$ are outliers that do not follow the Gumbel distribution while other $N - s$ data are sampled from the same Gumbel distribution. Note that, unlike our notation, N data were sampled from the population distribution and number of all data was $N + s$ in Ueda [1996, 2009]. Then, AIC with the s outliers is represented as

$$\frac{1}{2}\mathrm{AIC}_s = -\sum_{j=s+1}^{N} \log f(x_j \mid \theta') - \log(N-s)! + s \tag{4}$$

[Ueda, 1996, 2009; Marmolejo-Ramos et al., 2015], where f is the assumed PDF the samples follow, and θ' is the maximum likelihood parameters. In the original method, f has been assumed to be the normal distribution [Ueda, 1996, 2009]. However, the original method tends to be sensitive and detect too many values as outliers if the true ditribution is positively skewed [Marmolejo-Ramos et al., 2015]. In our case, we assume that the true distribution is approximated by the Gumbel distribution that has positive skewness. Therefore, instead of the normal distribution, $f(x_j \mid \theta) = P_G(x_j \mid \mu', \sigma')$ should be considered, where $P_\mathrm{G} := \frac{dF_\mathrm{G}}{dx}$ is the PDF of the Gumbel distribution.

In the following, we do not directly calculate eq.(4) that contains uncalculatable huge number $\log(N-s)!$ for our case ($N \sim 10^6$). Instead, for sufficiently large N, the difference in AIC between the cases of s outliers and $s+1$ outliers, $\frac{1}{2}d\mathrm{AIC}_s$,

can be approximated as

$$\begin{aligned}\frac{1}{2}d\mathrm{AIC}_s :=& \frac{1}{2}\left(\mathrm{AIC}_{s+1} - \mathrm{AIC}_s\right) \\ &\sim \log P_{\mathrm{G}}\left(x_{s+1} \mid \mu', \sigma'\right) + \log\left(N - s\right) + 1.\end{aligned} \tag{5}$$

Strictly, the maximum likelihood parameters based on all N data could differ from those estimated using $N - s$ or $N - s - 1$ data. However, we assume that $N \gg s$ holds and the parameters do not change significantly after the elimination of s data; see also Appendix B on its effect. Because we focus on the right tail of P_{G} and x_i is in the descending order, $P_{\mathrm{G}}(x_s) < P_{\mathrm{G}}(x_{s+1})$ holds, which results in

$$\frac{1}{2}d\mathrm{AIC}_s < \frac{1}{2}d\mathrm{AIC}_{s+1}. \tag{6}$$

In other words, the difference in AIC is a monotonically increasing sequence. If $d\mathrm{AIC}_s < 0$ holds, from the definition, we can reasonably regard that x_{s+1} is also an outlier rather than a sample from the Gumbel distribution. On the contrary, if $d\mathrm{AIC}_s > 0$ holds, the monotonicity guarantees that the difference is always positive as s increases. Thus, all x_i $(i > s)$ are not outliers. Finally, our procedure schematically illustrated in Figure 2 is as follows. We first obtain the MLE of the parameters μ' and σ', and then calculate $d\mathrm{AIC}_s$ for $s = 0, 1, 2, \ldots$. We stop the calculation when s reaches s_0, which makes $d\mathrm{AIC}_s$ positive for the first time, and finally conclude that $x_1, x_2, \ldots, x_{s_0}$ are outliers.

3 Case Study

3.1 Data

We considered a foreshock sequence of an M5.4 earthquake: origin time = 2011-06-30 08:16:37:06(JST); epicenter = 35.188N, 137.955E; depth = 4.3 km. According to the JMA catalog, 27 small foreshocks were recorded within 13 h before the mainshock (Table S.1); their epicenters are within 1 km from the epicenter of mainshock and surrounded by 4 Hi-net observation stations within 10 km (Figure 3), which may enable us to investigate the seismicity precisely. Thus, for each station and component, the 27 template waveforms were extracted from 0.5 s before each arrival of P-wave, and their length was 5 s (= 500 samples), such that the significant part of S-wave and its coda are included. To search events similar to these foreshocks, a criterion for outlier detection based on the empirical distribution of CC is required. We thus calculated the Network Cross-correlation Coefficient (NCC)

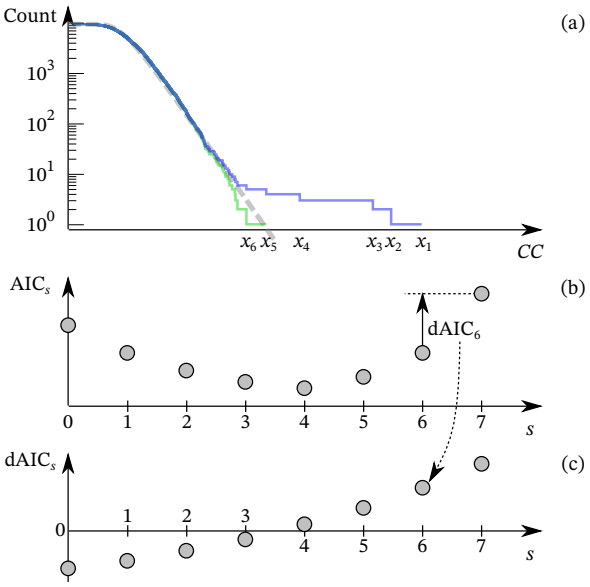

Figure 2. Schematic illustration for estimating $s_0 = 4$, where s_0 is the number of outliers ouf of $N = 10^4$. (a) Cumulative number of raw data (blue steps), estimated Gumbel distribution (gray broken line), and cumulative number of data after elimination of $x_1, \ldots, x_{s_0}$ (green steps). (b) Dependence of AIC on the number of outlier candidates, s. (c) Dependence of $d\mathrm{AIC}_s := \mathrm{AIC}_{s+1} - \mathrm{AIC}_s$ on s, where the definition is exemplified for $s = 6$. Even though the blue step due to x_5 is above the gray line in (a), x_5 is not regarded as the outlier because the step becomes closer to the gray line after the elimination.

among template waveforms due to the 27 events and 2-years continuous waveforms between 2009-06-29 and 2011-06-28 before the activation of the foreshocks. NCC is the averaged value of CC obtained in each station and component after shifting CC by lags between the origin time and arrival time of P-wave [Gibbons & Ringdal, 2006]. Even after averaging, maxima of NCC should follow GEV because maxima generated by arbitrary distribution follow GEV [Gumbel, 1958; Coles, 2001]. In our case, we stack 12 CC time series based on 3 components of the 4 stations and obtain 27 histograms of NCC in total. Before the calculation, we applied a band-pass filter to focus on the frequency band, in which waveforms due to foreshocks show high S/N ratios. Although Doi & Kawakata [2012] applied a band-pass filter of 15–40 Hz, we applied the filter of 5–30 Hz depending on the spectra of template waveforms; some automatic and objective determination method of the band should be developed in the future. We eliminated 15-s daily data of 09:00:00.00-09:00:15.00 for

checking the state-of-health of the observation instrument of Hi-net to ensure that CC was not affected by the test signals.

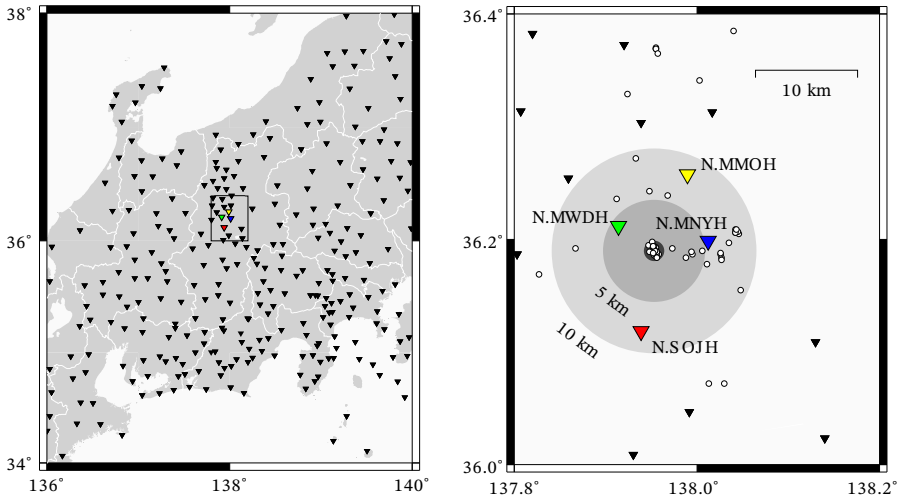

Figure 3. Distribution of observation points (triangles) and foreshock hypocenters (white circles) prior to an M5.4 mainshock in Nagano, Japan. Waveforms observed at N.MWDH (green), N.MNYH (blue), N.MMOH (yellow), and N.SOJH (red) stations were analyzed in this study. See Table S.1 for detail of the 27 events within 1 km from the epicenter shown in the darkest circle.

3.2 Result: Histogram of NCC

All histograms of NCC are shown in Figure 4. The histograms were normalized by their standard deviation, which means that they should be well approximated by the standard normal distribution plotted by the gray parabola in the semi-log plot if NCC follows a normal distribution. However, the tails of the NCC histograms appear to be linear in the semi-log plot, and significantly different from the theoretical distribution discussed in the previous section. The difference between the theoretical model and empirical data is over a hundredfold in 7σ and ten thousandfold in 8σ. Therefore, the possibility of false positives may be severely underestimated if we set the threshold value as 8σ [Aso et al., 2011; Kato et al., 2012] and implicitly assume that the histogram follows a normal distribution. This fact strongly implies that the observed microtremor is significantly far from the i.i.d. assumed in the ideal case and has non-negligible self-interdependence. Weekly statistics of NCC histograms (Figure. 5) show that the standard deviation is higher than the case of the random

waveform ($\sigma = \sqrt{1.8\,d^{-1}}$, where $d = 500 \times 12$ in this case), which implies that the microtremor is somehow biased. Hence, we should refer to the distribution of the maxima of NCC that is less sensitive to the self-interdependence. Figure 5 also shows that characteristics of histograms, such as the standard deviation and kurtosis, fluctuated immediately after the week, including those on March 11, 2011, the day the M9.0 Tohoku earthquake occurred. However, the 2 years were not separated in our analysis because a sufficient amount of data is required to investigate the tails of histograms.

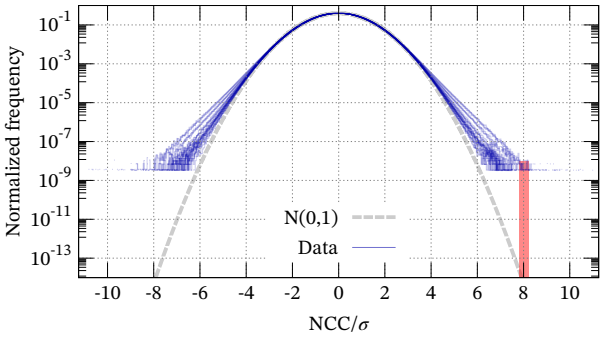

Figure 4. Empirical distribution of NCC between 2-years continuous records and 27 template waveforms (blue). The abscissa is normalized by the standard deviation. Vertical red bar indicates that the empirical distribution is several orders larger than the normal distribution (gray) in 8σ.

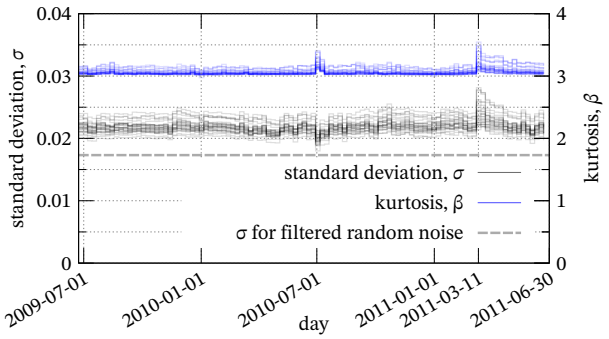

Figure 5. Temporal variation of the standard deviation and kurtosis of the empirical distribution of NCC.

3.3 Result: Cumulative Distribution of max. of NCC

We attempted to detect seismic events that possibly occurred in the 2 years using the proposed method in 2.3 after fitting the cumulative number of the maxima of NCC at every minute between 2009-06-29 and 2011-06-28. In total, we could select 21 outliers according to Figure 6, which shows the cumulative number of calculated NCC, the estimated Gumbel distribution $1 - F_{\mathrm{G}}\left(x \mid \mu', \sigma'\right)$, and detected outliers. However, we classified some of these outliers as the same event because they emerged within 1 s. Finally, we could detect 4 new events, as shown in Table 1, which have not been cataloged by the JMA. As shown in Figure S.1–S.4, the detected waveforms show amplitudes of maximum 10–20% of the template waveforms and, therefore, have relatively low S/N ratios compared to the template. Even from such noisy data, our method provided the seismic signals without any prescribed threshold. The finding of the triplet similar events 3–4 days before the mainshock in the foreshock region (IDs B–D in Table 1) may provide us with new insight for considering the preparation process of the mainshock.

Table 1. Detected events by the proposed algorithm.

ID	date	time	similar to
A	2011-05-04	19:17:00	05, 06
B	2011-06-26	11:57:47	01, 02, 04, 14, 18, 27
C	2011-06-26	12:57:45	01, 02, 04, 18, 23, 27
D	2011-06-27	07:24:14	01, 02, 04, 18, 20, 23, 27

4 Discussion

Compared to conventional thresholding methods, the most important advantage of our method is that the results are objective and reasonable; the result is less affected by arbitrariness in principle. We can suggest the possibility of false positives under the Gumbel distribution because the differences between the distribution and cumulative number of data are almost less than tenfold (Figure 6). The conventional method involves a trade-off between the number of detected events and false positive depending on the threshold value. In our method, however, the detection criterion is automatically determined depending only on the quality of data. Only 4 events were

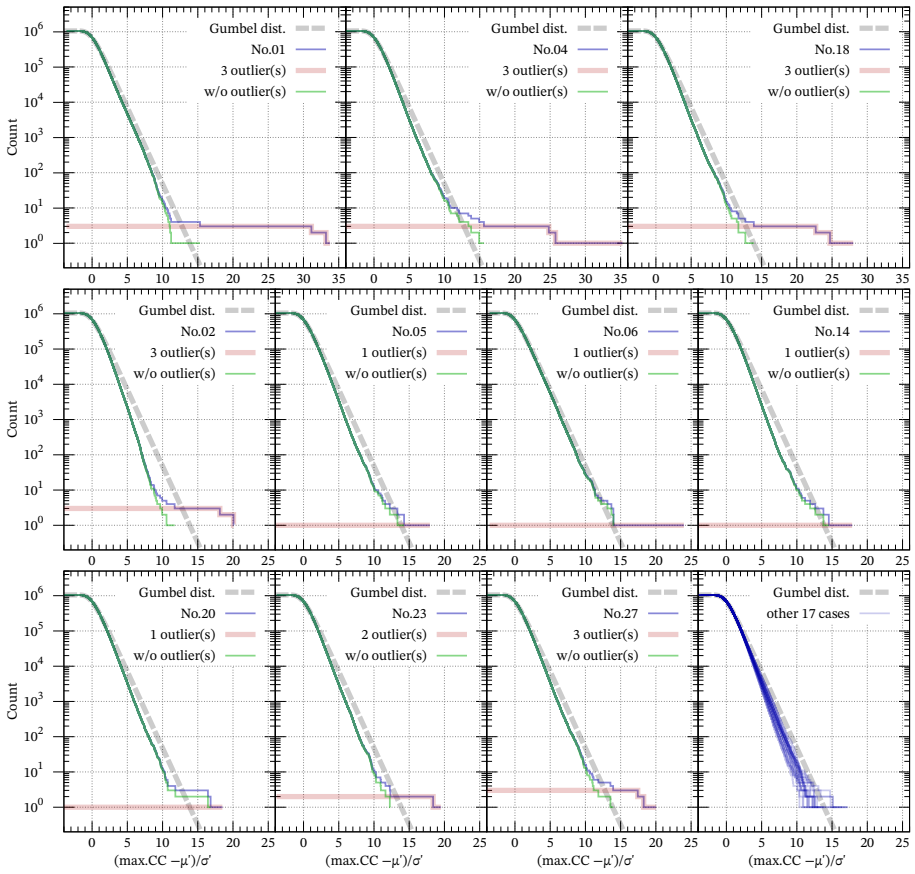

Figure 6. Cumulative distributions of normalized NCC (blue) and the Gumbel distribution with $\mu' = 0$ and $\sigma' = 1$ as the fitting curve (gray) for 27 templates of Table S.1. Red steps indicate detected outliers in terms of the minimization of AIC. 17 cases accompanied by no outliers are plotted all together in the right bottom.

detected in our analysis, which may mean that the hypocenter region had been quite inactive before the foreshock activity or our method is excessively strict at finding many uncataloged events. Nevertheless, even if the latter is true, the detection of 4 uncataloged events shows that our method has higher detection ability than that of JMA at that term, at least for similar seismic events.

It is noteworthy that our method is not completely free of arbitrariness. One concern is the length of intervals using which we selected the maxima. In our experiment, we selected an interval of 1 min (i.e., 6,000 samples) considering computational time, but in principle, the interval can be, for example, 5 seconds (i.e., 500 samples). With longer lengths, the data distribution may converge to the Gum-

bel distribution [Gumbel, 1958; Coles, 2001], but the temporal resolution will decrease because relatively smaller peaks of CC values will be neglected if a higher peak emerges in the same interval, which becomes likely for longer intervals. In contrast, with shorter lengths, the convergence might not be achieved, and the data will require fitting using not the Gumbel (3) but the GEV (2) distribution, which includes one more parameter and is time-consuming (Appendix C). Therefore, the effect of the length should be quantified theoretically and practically in the future.

The background level of CC may have daily variation [Aso et al., 2011] or long-term variation as shown in Figure 5, and, for precise analysis, the threshold should be determined in each term (e.g., diurnal and nocturnal distribution of CC). In such a case, our method can be applied to each term separately, although we ignored such variations for simplicity.

Because we analyzed continuous records only from 4 stations, it remains unclear whether the empirical distribution can be modeled by the Gumbel distribution in general cases. A suitable approximation is possible using other limits of GEV: the Fréchet or Weibull distribution. In practice, the shape of the tail should be further investigated considering these possibilities in each analysis.

5 Conclusion

We developed an objective matched-filter technique based on AIC and the extreme value theory. We showed that the CC between any template and i.i.d. random waveform follows the normal distribution, which provides a reference for examining the deviation of data from the i.i.d. case. To reduce the possibility of a false positive, we considered the maximum of CC in each interval and found that the maxima follows the Gumbel distribution. Finally, using the distribution and AIC, we propose a reasonable method for detecting outlier seismic signals that is less sensitive to arbitrariness than a conventional thresholding method. Regardless of whether NCC follows the normal distribution, the proposed method can be applied to analyses of seismic event detection.

Acknowledgments

We used continuous waveform records of the NIED high sensitivity seismograph network of Japan (https://doi.org/10.17598/NIED.0003) and JMA unified earthquake catalog(https://www.hinet.bosai.go.jp/REGS/JMA/jmalist.php?LANG=en). This work was supported by MEXT, under its Earthquake and Volcano Hazards Observation and Research Program. Yutaka Toyomoto contributed to waveform picking and preliminary calculations. We thank Kazuaki Ohta, Naofumi Aso, Hiroyuki Noda and Toru Imai for helpful discussions.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, in *2nd International Symposium on Information Theory* pp.267–281, Akademiai Kiado.

Aso, N., Ohta, K., & Ide, S. (2011). Volcanic-like low-frequency earthquakes beneath Osaka Bay in the absence of a volcano. *Geophys. Res. Lett.*, **38**(8), L08303. http://doi.org/10.1029/2011GL046935

Aso, N., Ohta, K., & Ide, S. (2013). Tectonic, volcanic, and semi-volcanic deep low-frequency earthquakes in western Japan. *Tectonophysics*, **600**, 27–40. http://doi.org/10.1016/j.tecto.2012.12.015.

Bouchon, M., Karabulut, H., Aktar, M., Özalaybey, S., Schmittbuhl, J., & Bouin, M. P. (2011). Extended nucleation of the 1999 Mw 7.6 Izmit earthquake. *Science*, **331**(6019), 877–880. http://doi.org/10.1126/science.1197341

Coles, S., (2001). *An introduction to statistical modeling of extreme values*, Springer-Verlag, London.

David, H. A., & Nagaraja, H. N. (2003). *Order statistics*, 3rd edn. Wiley Series in Probability and Statistics, John Wiley & Sons, New Jersey.

Doi, I., & Kawakata, H. (2012). A non-accelerating foreshock sequence followed by a short period of quiescence for a large inland earthquake. *Geophys. Res. Lett.*, **39**(11). http://doi.org/10.1029/2012GL051779

Doi, I., & Kawakata, H. (2013). Spatio-temporal occurrence patterns among the foreshocks preceding the 2007 Noto Hanto earthquake. *Earth Planets Space*, **65**(9), 1053–1058. http://doi.org/10.5047/eps.2013.04.001

Gibbons, S. J., & Ringdal, F. (2006). The detection of low magnitude seismic events using array-based waveform correlation. *Geophys. J. Int.*, **165**(1), 149–166. http://doi.org/10.1111/j.1365-246X.2006.02865.x

Gumbel, E. J. (1958). *Statistics of Extremes*, Dover.

Hosking, J. R. M. (1990). L-moments: analysis and estimation of distributions using linear combinations of order statistics, *J. Roy. Stat. Soc. B.* **52**(1) 105–124. https://doi.org/10.1111/j.2517-6161.1990.tb01775.x

Kadota, K., Nishimura, S.-I., Bono, H., Nakamura, S., Hayashizaki, Y., Okazaki, Y., & Takahashi, K. (2003). Detection of genes with tissue-specific expression patterns using Akaike's information criterion procedure. *Physiol. Genomics*, **12**(3), 251–259. http://doi.org/10.1152/physiolgenomics.00153.2002

Kato, A., Obara, K., Igarashi, T., Tsuruoka, H., Nakagawa, S., & Hirata, N. (2012). Propagation of slow slip leading up to the 2011 Mw 9.0 Tohoku-Oki earthquake, *Science*, **335**, 705–708. http://doi.org/10.1126/science.1215141

Kitagawa, G. (1979). On the use of AIC for the detection of outliers. *Technometrics*, **21**(2), 193–199. http://doi.org/10.1080/00401706.1979.10489749

Marmolejo-Ramos, F., Vlez, J.I., & Romo, X. (2015). Automatic detection of discordant outliers via the Ueda's method, *J. Stat. Distrib. App.*, **2**:8. https://doi.org/10.1186/s40488-015-0031-y

Martins, E. S., & Stedinger, J. R. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data, *Water Resour. Res.*, **36**(3) 737–744. https://doi.org/10.1029/1999WR900330

Maxwell, J. C. (1860). V. Illustrations of the dynamical theory of gases. –Part I. On the motions and collisions of perfectly elastic spheres, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 19(124), 19–32. https://doi.org/10.1080/14786446008642818

Ohmi, S. (2015). Seismic activity near Mt. Hotaka in Hida mountain range in 2013, detected by the matched filter method. *Zisin, Series 2*, **68**(1), 1–15. (in Japanese with English abstract) https://doi.org/10.4294/zisin.68.1

Ohta, K., & Ide, S. (2008). A precise hypocenter determination method using network correlation coefficients and its application to deep low-frequency earthquakes. *Earth Planets Space*, **60**(8), 877–882. http://doi.org/10.1186/BF03352840

Shelly, D. R., Beroza, G. C., & Ide, S. (2007). Non-volcanic tremor and low-frequency earthquake swarms. *Nature*, **446**(7133), 305–307. http://doi.org/10.1038/nature05666

Shimojo, K., Enescu, B., Yagi, Y., & Takeda, T. (2014). Fluid-driven seismicity activation in northern Nagano region after the 2011 M9.0 Tohoku-oki earthquake. *Geophys. Res. Lett.*, **41**(21), 7524–7531. http://doi.org/10.1002/2014GL061763

Squire, W., & Trapp, G. (1998). Using Complex Variables to Estimate Derivatives of Real Functions, *SIAM Rev.*, **40**(1) 110–112. https://doi.org/10.1137/S003614459631241X

Ueda, T. (1996). Simple method for the detection of outliers. *Japanese J. Appl. Stat.*, **25**(1), 17-26 (in Japansese). https://doi.org/10.5023/jappstat.25.17

Ueda, T. (2009). A simple method for the detection of outliers. *Electron. J. App. Stat. Anal.*, **1**, 67–76. (F. Marmolejo-Ramos & S. Kinoshita, Trans.) (Original work published in 1996) http://doi.org/10.1285/i20705948v2n1p67

Appendix A Approximation of the CC distribution

Here, we show that the normal distribution $\mathcal{N}\left(0, d^{-1}\right)$ approximates the theoretical distribution of CC between d-dimensional two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ extracted from a continuous waveform record. First, we show that the extracted vector is statistically isotropic. From the definition, we consider $\boldsymbol{v}^{(t)} = (x_t, x_{t+1}, \ldots, x_{t+d-1})$, where x_t is the t-th component of continuous record. Therefore, $CC_t = \widehat{\boldsymbol{u}} \cdot \widehat{\boldsymbol{v}}^{(t)}$ is the t-th value of CC if $\boldsymbol{u}$ is the fixed template. If $t_0 \in [t, t+d-1]$ exists such that $|x_{t_0}|$ is significantly larger (or smaller) than others, $\widehat{\boldsymbol{v}}^{(t)}$ itself is strongly (or less) oriented to the t_0-th direction. However, simultaneously, $\widehat{\boldsymbol{v}}^{(t+1)}$, $\widehat{\boldsymbol{v}}^{(t+2)}$, $\widehat{\boldsymbol{v}}^{(t+3)}\ldots$ are strongly (or less) oriented towards the $t_0 - 1$, $t_0 - 2$, $t_0 - 3\ldots$ direction; this discussion is obviously valid even if the continuous record has some self-interdependency. Therefore, it is impossible to give some tendency to the direction of $\boldsymbol{v}$, that is, $\boldsymbol{v}^{(t)}$ for all t is statistically isotropic.

Given the isotropy, the normal distribution $\mathcal{N}\left(0, d^{-1}\right)$ can be obtained as the extension of the derivation of the Maxwell-Boltzmann distribution. However, Maxwell (1860) assumed that each component of the vector, which is 3-dimensional and d-dimensional in the original and our problem, respectively, is independent; in our problem, this assumption does not hold because of $|\widehat{\boldsymbol{u}}|^2 = |\widehat{\boldsymbol{v}}|^2 = 1$, where $\widehat{\boldsymbol{u}}$ and

$\widehat{\boldsymbol{v}}$ are normalized $\boldsymbol{u}$ and $\boldsymbol{v}$, respectively, after elimination of their offset. Therefore, we loosen this constraint as $E\left(|\widehat{\boldsymbol{v}}|^2\right) = 1$, where $E(\cdot)$ indicates the mean value. After the derivation, we justify this assumption for larger value of d.

The derivation of the Maxwell-Boltzmann distribution is purely mathematical rather than physical. Maxwell [1860] considered that each component of particle velocity $\boldsymbol{v} = (v_1, \ldots, v_d)$ is a random variable that follows the same PDF, P. Although only $d = 3$ was considered in the original, we extend it into the general case. In the following, we consider the PDF for each component of $\widehat{\boldsymbol{v}} = (\widehat{v}_1, \ldots, \widehat{v}_d)$.

By assuming that the random vector $\boldsymbol{v}$ is statistically isotropic (i.e., "*the directions of the coordinates are perfectly arbitrary*", Maxwell wrote), $E(\widehat{v}_j) = 0$ holds for arbitrary direction, and the joint probability of $\widehat{v}_1, \ldots, \widehat{v}_d$ is coordinate-free and depends only on $|\widehat{\boldsymbol{v}}|^2$ written as

$$\prod_{j=1}^{d} P\left(\widehat{v}_j\right) = \phi\left(\sum_{j=1}^{d} \widehat{v}_j^2\right). \tag{A.1}$$

Because only an exponential function satisfies this property,

$$P\left(\widehat{v}_j\right) = \frac{1}{\alpha\sqrt{\pi}} \exp\left(-\frac{\widehat{v}_j^2}{\alpha^2}\right) \tag{A.2}$$

is obtained under the condition of $\int_{\mathbb{R}} P(x)dx = 1$, where α is a positive parameter to be determined. The joint probability is written as

$$\prod_{j=1}^{d} P\left(\widehat{v}_j\right) = \frac{1}{\alpha^d \pi^{d/2}} \exp\left(-\frac{|\widehat{\boldsymbol{v}}|^2}{\alpha^2}\right), \tag{A.3}$$

and the mean value of $|\widehat{\boldsymbol{v}}|^2$ is

$$E\left(|\widehat{\boldsymbol{v}}|^2\right) = \int_{\mathbb{R}^d} |\widehat{\boldsymbol{v}}|^2 \prod_{j=1}^{d} P\left(\widehat{v}_j\right) d\widehat{v}_j \tag{A.4}$$

$$= \frac{1}{\alpha^d \pi^{d/2}} \int_{S_{d-1}} d\omega \int_0^\infty |\widehat{\boldsymbol{v}}|^2 \, |\widehat{\boldsymbol{v}}|^{d-1} \exp\left(-\frac{|\widehat{\boldsymbol{v}}|}{\alpha^2}\right) d\,|\widehat{\boldsymbol{v}}| \tag{A.5}$$

$$= \frac{2\pi^{d/2}}{\alpha^d \pi^{d/2} \Gamma\left(d/2\right)} \frac{1}{2} \alpha^{d+2} \Gamma\left(\frac{d}{2} + 1\right) \tag{A.6}$$

$$= \frac{\alpha^2 d}{2}, \tag{A.7}$$

where $S_{d-1} = 2\pi^{d/2}/\Gamma(d/2)$ is the area of $(d-1)$-dimensional unit sphere, $d\omega$ is the solid angle, and Γ is the Gamma function. $|\widehat{\boldsymbol{v}}|^{d-1}$ is derived from the Jacobian, and we use

$$\int_0^\infty x^p \exp\left(-\frac{x^2}{a^2}\right) dx = \frac{1}{2} a^{p+1} \Gamma\left(\frac{p+1}{2}\right). \tag{A.8}$$

Finally, with $E\left(|\widehat{\boldsymbol{v}}|^2\right) = 1$, we get

$$\alpha^2 = \frac{2}{d}, \tag{A.9}$$

which yields

$$P(\widehat{v}_j) = \sqrt{\frac{d}{2\pi}} \exp\left(-\frac{\widehat{v}_j^2}{2d^{-1}}\right). \tag{A.10}$$

Obviously, if d is small, eq.(A.10) does not approximate the distribution of CC because the probability is not negligible for $|\widehat{v}_j| > 1$. Hence we have to consider the sufficiently large value of d that makes $P(|\widehat{v}_j| > 1)$ negligibly small. Moreover, the variance of $|\widehat{\boldsymbol{v}}|^2$ calculated as

$$E\left(\left(|\widehat{\boldsymbol{v}}|^2 - 1\right)^2\right) = E\left(|\widehat{\boldsymbol{v}}|^4\right) - 2E\left(|\widehat{\boldsymbol{v}}|^2\right) + E(1) \tag{A.11}$$

$$= \alpha^4 \left(\frac{d}{2} + 1\right) \frac{d}{2} - 2 + 1 \tag{A.12}$$

$$= \frac{2}{d}, \tag{A.13}$$

means that the possibility of $|\widehat{\boldsymbol{v}}|^2 = 1$ in the strict sense becomes larger as d increases. Therefore, the constraint $|\widehat{\boldsymbol{v}}|^2 = 1$ is approximately satisfied for larger values of d.

Considering that both Maxwell's particle and our unit random vector is isotropic, the PDF (A.10) provides not only the specific component $\widehat{v}_j$ but also a component along all directions including $\widehat{\boldsymbol{u}}$ in the same manner. Therefore, the inner product of an arbitrary random unit vector $\widehat{\boldsymbol{v}}$ extracted from random continuous waveform and arbitrary fixed unit vector $\widehat{\boldsymbol{u}}$ approximately follows the normal distribution with the variance of d^{-1}.

Appendix B Does not the log-likelihood vary significantly after elimination of some data?

In eq.(5), we assume that the maximum likelihood parameters for all N, $N - s$ and $N - s - 1$ data do not vary significantly because $N \gg s$ holds. Even so, the effect of the small difference of the parameters on AIC_s appears to be unclear. In the calculation of the log-likelihood, $\sum_{j=s+1}^{N} \log P(x_j \mid \mu', \sigma')$, even negligibly small difference could be stacked and possibly become a significant amount.

However, we can show that the stacked amount is still negligible. Let the parameters (μ', σ') and (μ'', σ'') be the MLE using all N data and $N - s$ data, respec-

tively. Therefore, the error of AIC (i.e., error of the log-likelihood) using the former instead of latter has the same order of the Kullback–Leibler divergence

$$D(P_G(\mu'', \sigma''), P_G(\mu', \sigma')) = \int_{\mathbb{R}} P_G(x \mid \mu'', \sigma'') \log \frac{P_G(x \mid \mu'', \sigma'')}{P_G(x \mid \mu', \sigma')} dx. \tag{B.1}$$

This is equivalent to the loss function defined in eq.(3.1) of Akaike [1973], and depends only on the second or higher order of $(\mu'' - \mu')$ and $(\sigma'' - \sigma')$ after the Taylor series expansion; see eq.(4.5) of Akaike [1973]. Hence, any small error of $(\mu'' - \mu')$ or $(\sigma'' - \sigma')$ does not vary the log-likelihood significantly.

Appendix C MLE of GEV parameters

MLE of GEV parameters is equivalent to solving the equations below with respect to μ', σ', and k:

$$\begin{aligned} \sum_{i=1}^{N} \frac{z_i}{y_i} &= 0, \\ -N + \sum_{i=1}^{N} \frac{z_i}{y_i} \left(\frac{x_i - \mu'}{\sigma'} \right) &= 0, \\ \sum_{i=1}^{N} \left(z_i \log(y_i) + \frac{z_i}{y_i} \left(\frac{x_i - \mu'}{\sigma'} \right) \right) &= 0 \end{aligned} \tag{C.1}$$

[Martins & Stedinger, 2000; Coles, 2001], where $y_i := 1 + (k/\sigma')(x_i - \mu')$ and $z_i := 1 + k - y_i^{-1/k}$ (note: k is opposite in sign between Coles [2001] and Martins & Stedinger [2000]), and we eliminate some unnecessary coefficients. To solve them using the Newton-Raphson method, the Hessian matrix that is the derivative of eqs.(C.1) with respect to the 3 parameters should be calculated. Although the representations of the derivatives are slightly complicated, we simply compute the matrix by the automatic differentiation using a small complex variable [Squire & Trapp, 1998]. The initial values for iteration are given by L-moments [Hosking, 1990].

Unfortunately, the Newton-Raphson method sometimes fails during its iteration due to the following reason. During the MLE process, we have to calculate the log-likelihood $\log P_{\mathrm{GEV}}(x_j \mid \mu', \sigma', k)$ for all the sample x_j, where μ', σ', k is not necessarily the MLE of the parameters, which is attributable to the iteration. In case of $k < 0$, as mentioned in the main text, the PDF, P_{GEV}, for $x > \mu - \sigma'/k$ is zero. Therefore, we may substitute zero into P_{GEV} if $x_j > \mu - \sigma'/k$ holds, and the iteration stops due to the numerical error ($\log 0 = -\infty$). In particular, this error tends to

occur if the sample includes outliers, which is abnormally large. Hence, the MLE of GEV parameters is technically difficult, and we may require some ad hoc implementation.

Because we particularly focus on the case of the Gumbel distribution, the equations for maximum likelihood estimators are represented explicitly by taking $k \to \infty$:

$$\begin{aligned} \sum_{i=1}^{N} \left(1 - \exp\left(-\frac{x_i - \mu'}{\sigma'}\right)\right) &= 0, \\ -N + \sum_{i=1}^{N} \left(1 - \exp\left(-\frac{x_i - \mu'}{\sigma'}\right)\right)\left(\frac{x_i - \mu'}{\sigma'}\right) &= 0. \end{aligned} \tag{C.2}$$